SatCORPS Hybridized Cloud Product Data Storage: The Design of a Hybrid Data Repository That Leverages the Strengths of the Cloud and the Data Center

26th Conference on Satellite Meteorology, Oceanography, and Climatology
40th Conference on Environmental Information Processing Technologies
American Meteorological Society
104th Annual Meeting
January 28, 2024

Chee, T.[1], Nguyen, L.[2], Smith Jr., W. L. [2], Vakhnin, A., Palikonda, R.[3]
[1]*ADNet, Bethesda, MD* [2]*NASA Langley Research Center, Hampton, VA* [3]*Analytical Mechanics Associates, Hampton, VA*

Improvements in science = More storage

As instrument resolution, temporal range, data processing and scientific analysis techniques improve, larger datasets per unit are the result.

This information has to go somewhere and disk and robotic tape archive are the endpoints. Leveraging the strengths and weaknesses of the endpoints like cost, location, speed of response and redundancy can result in a reliable, resilient and manageable data repository.

This work describes the upcoming challenge and current architecture of the SatCORPS hybrid data repository that is geographically and storage type diverse.

The SatCORPS On-Site Data Repositor from the working side

Plot Source: NOAA Open Data Dissemination: Petabyte-scale Earth system data in the cloud, Willet, D., et al. Science Advances, 20 Sep. 2023, Vol 9, Issue 28, DOI: 10.1126/sciadv.adh0032

Our newest dataset is pushing us to new levels of data storage requirements

The Satellite ClOud and Radiative Property retrieval System (SatCORPS)

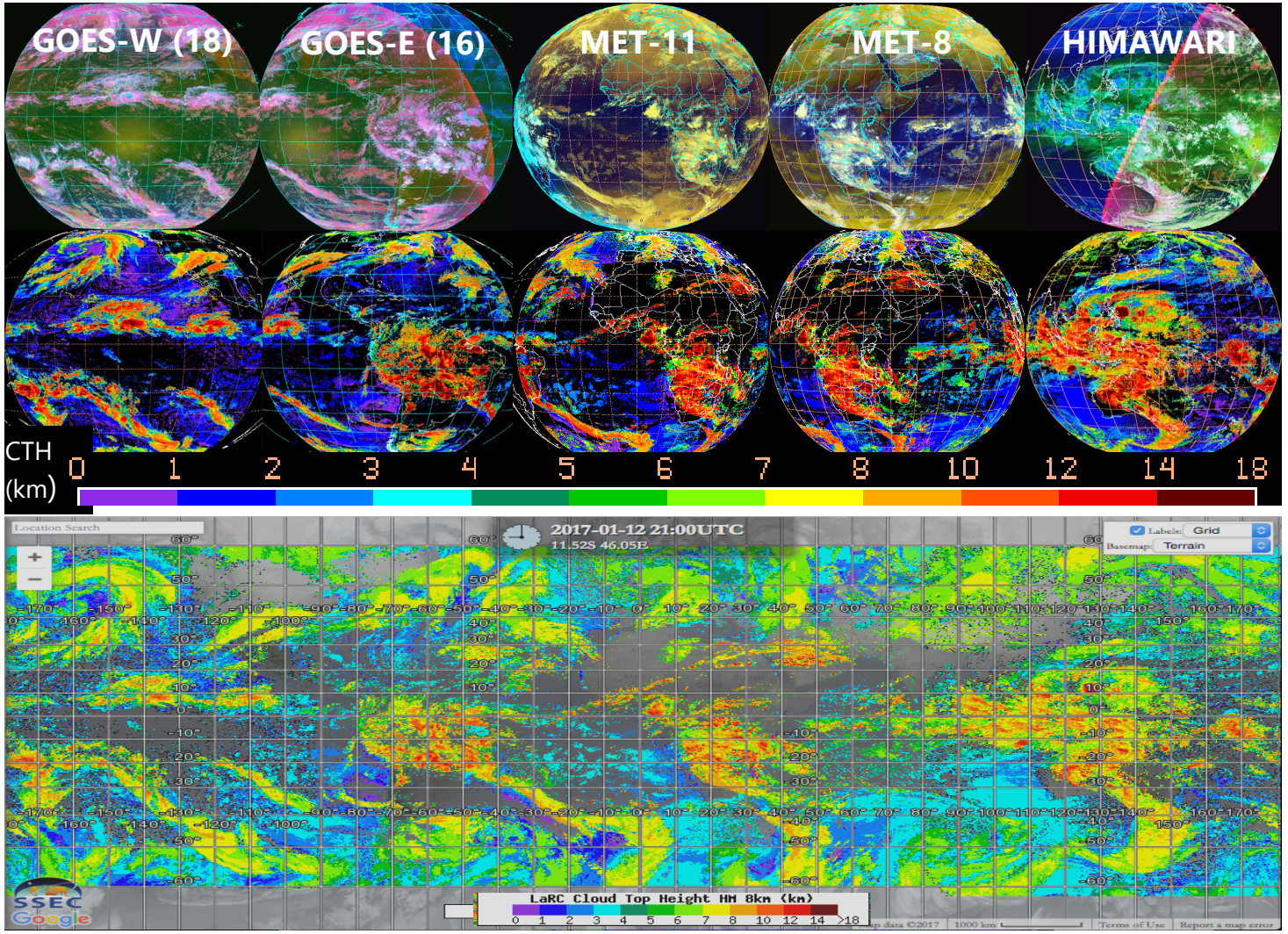
Unique SatCORPS Data Products

Traditional Standard Cloud Products

- Cloud Mask, Thermodynamic phase
- Cloud Temperature and Height
- Optical Depth, Effective Radius, Water Path
- Droplet number concentration

Innovative SatCORPS Data Products

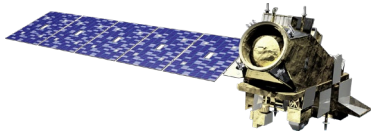
- Cloud optical properties at Night
  - Physical retrieval for thin cloud optical properties
  - Machine learning for thick clouds (diurnally consistent)
- Cloud vertical structure
  - Cloud layering, thickness, base heights
  - Cloud water content profiles
- Radiative Fluxes (TOA and SFC)
- Surface Skin Temperature
- Aviation weather hazards & climate impacts
  - Icing and convection
  - Contrail optical properties, radiative effects



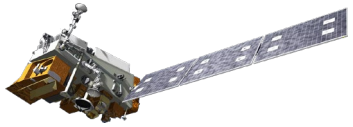
NRT hourly GEO Global Cloud Composite (GCC), 3km subsampled data products

The Low Earth Orbiting (LEO) Satellites Used in the SatCORPS GCC

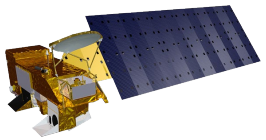
Low Earth Orbiting Satellites:



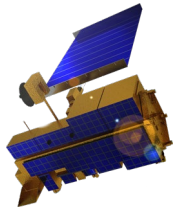
JPSS-2/NOAA-21
VIIRS / CrIS / IASI



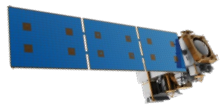
JPSS-1/NOAA-20
VIIRS / CrIS / IASI



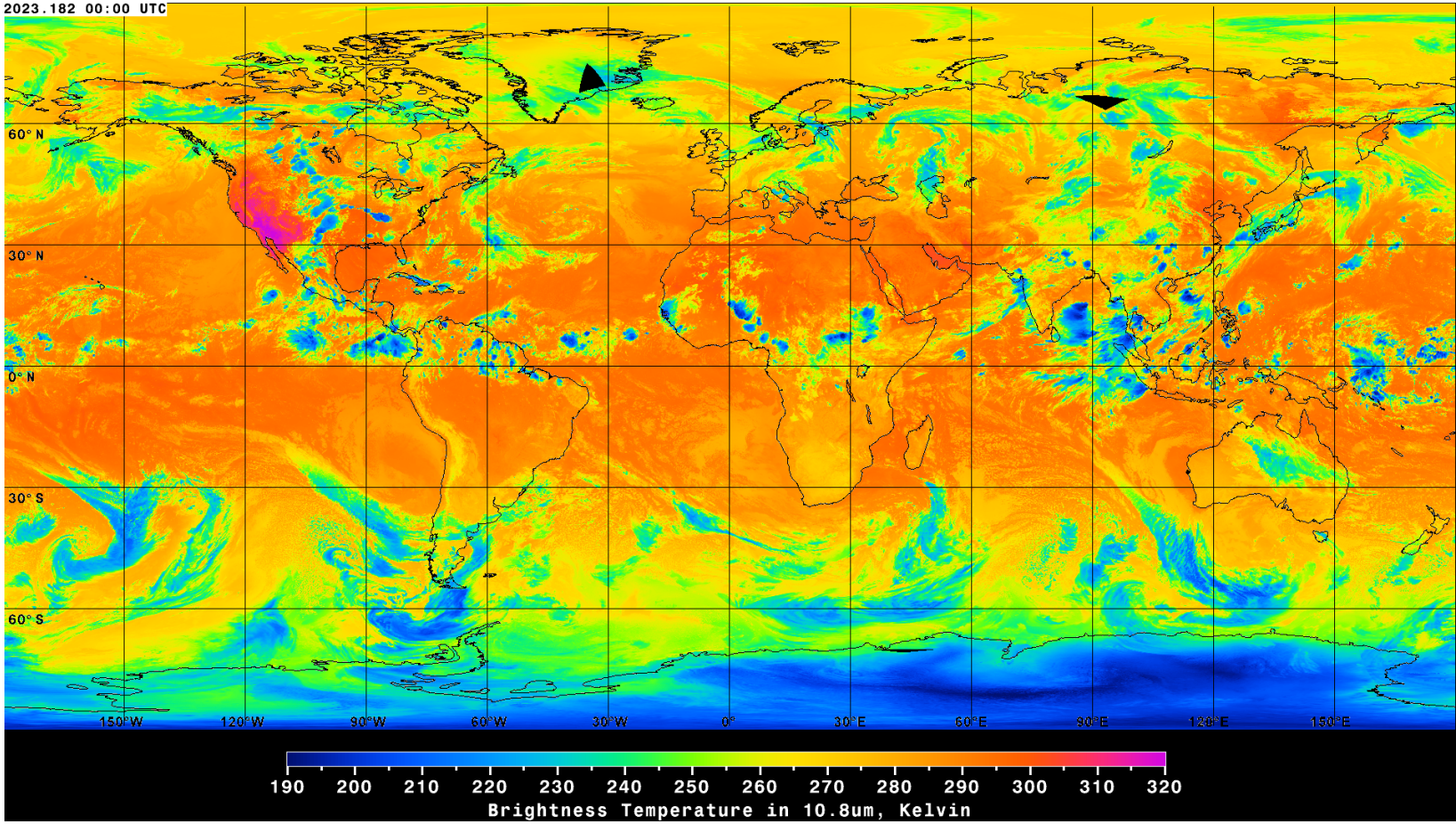
Aqua
MODIS



Terra
MODIS



Suomi NPP
VIIRS / CrIS / IASI



NRT hourly GEO Global Cloud Composite (GCC), 3km subsampled data products

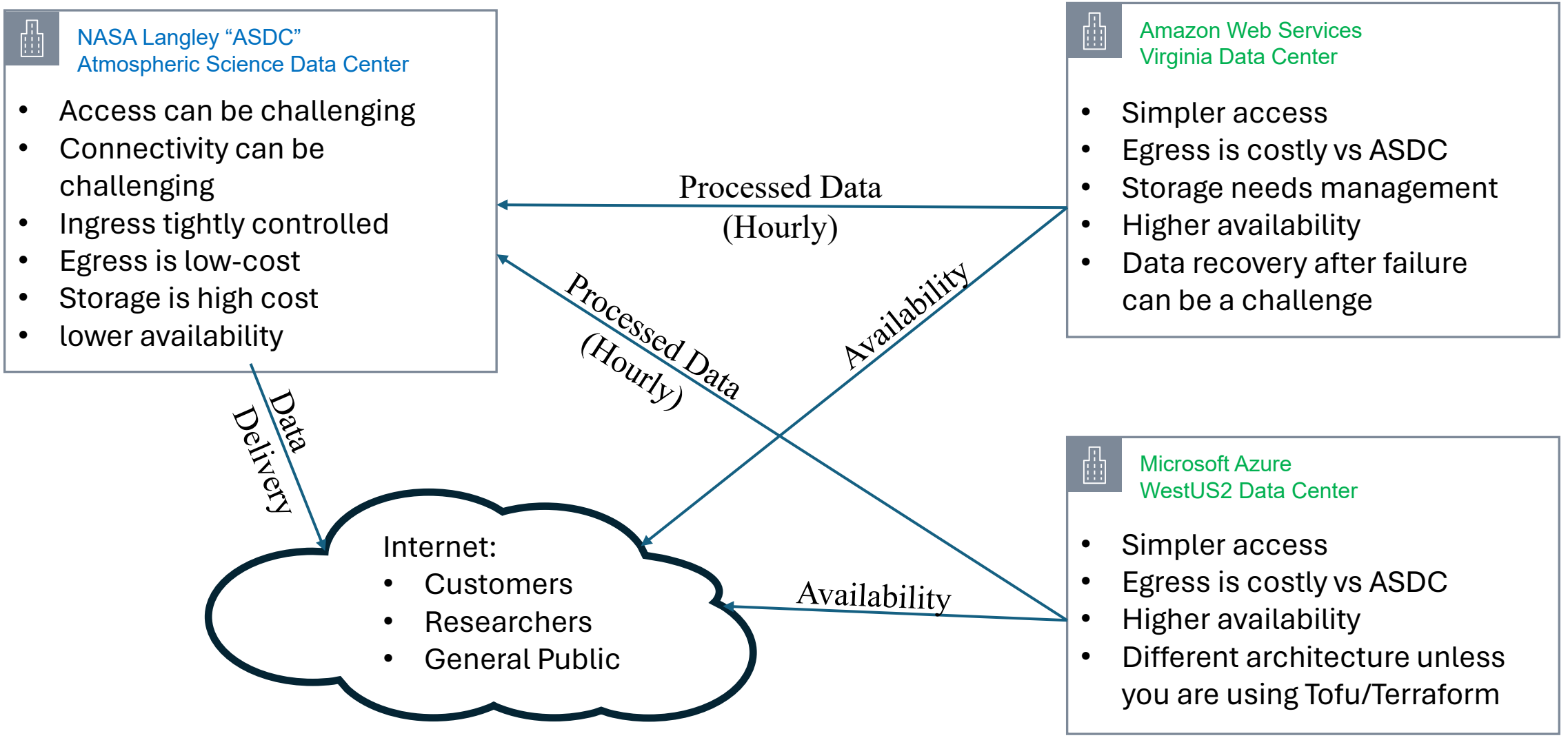
Data Repository Requirements

The requirements for our data repository have evolved over time as the system architecture has expanded:

- Retention of data – large size
- Scalable – room to grow
- Supports reliability
- Supports availability
- Fault Tolerant

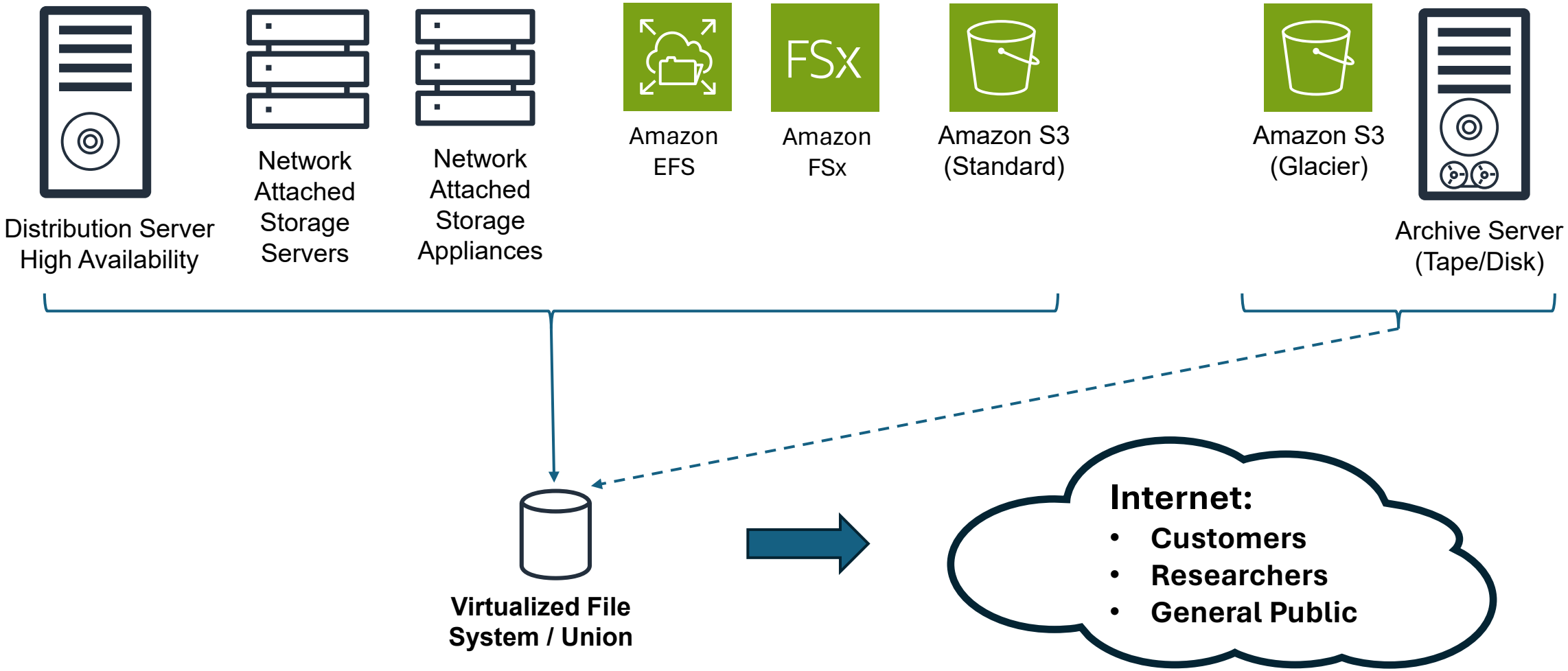
The LaRC Data Repository used by the SatCORPS Group

The Solution Employed: Hybridized Data Repository



A Hybridized Data Storage for Cloud Product Data

Different categories of storage with different attributes are used to create a data repository



Merging Technologies - Software

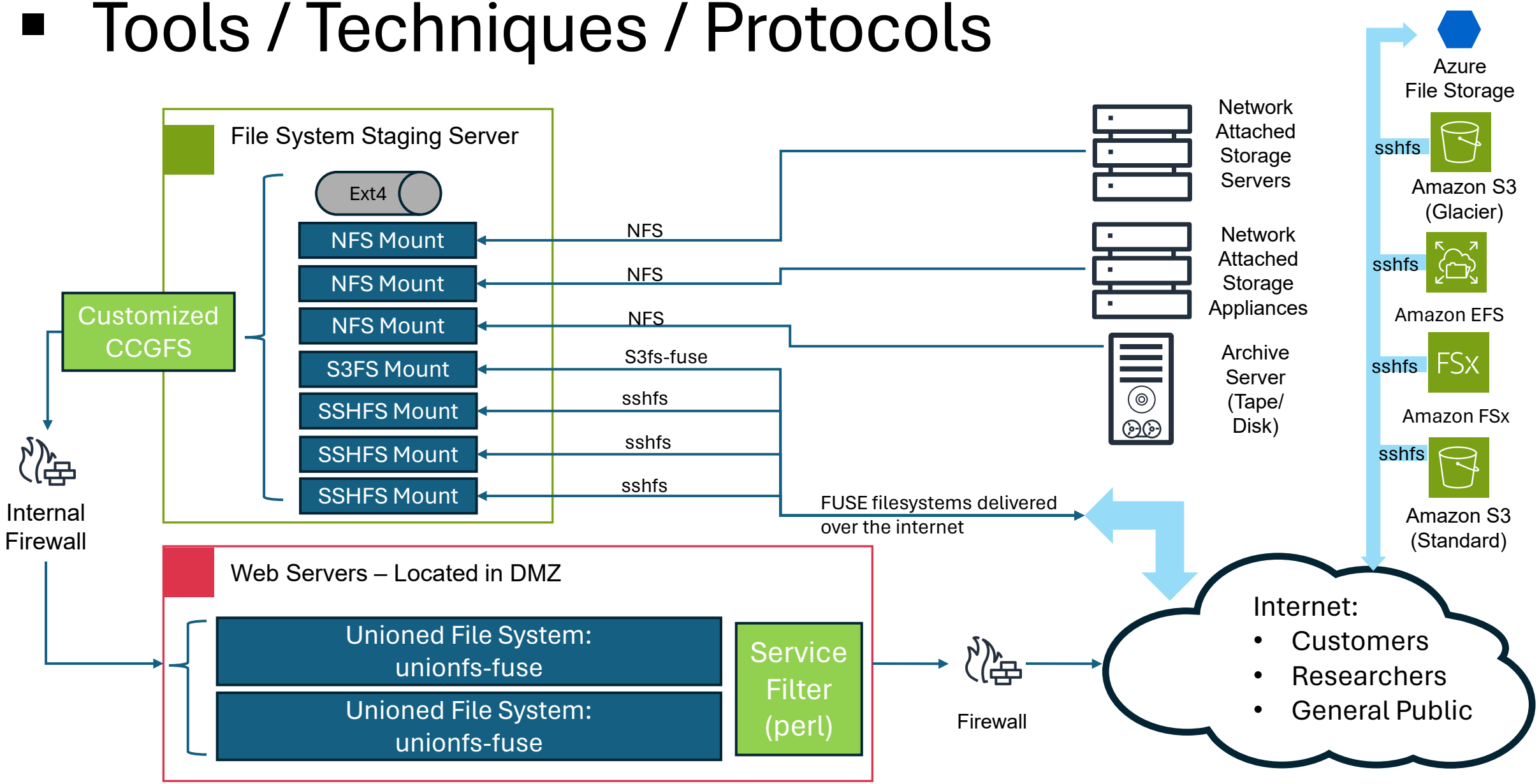
We use a number of different technologies to materialize the repository

- FUSE – FileSystem in Userspace *https://github.com/libfuse/libfuse*
  - CCGFS – CC Network Filesystem *https://inai.de/projects/ccgfs/*
  - S3FS: *https://github.com/s3fs-fuse*
  - Unionfs-fuse: *https://github.com/rpodgorny/unionfs-fuse*

- NFS – Network File System

Issues: Server Lock

Tools / Techniques / Protocols



Results / Findings

Hybridization has both pros and cons:

- maximizes functionality
- Improves redundancy
- Minimizes costs of delivery
- Adds complexity

Data Repository issues evolve over time necessitating maintenance processes:

- Evolution of solutions
- Documentation so stakeholders are aware of system state and path

Monitoring is crucial

- To control costs
- To identifying issues/bottlenecks